CDF Note 10625

Search for the Standard Model Higgs boson in $\tau^+\tau^-$ + jets final state with 8.3fb^{-1} of CDF data

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: August 26, 2011)

We present a search for the Standard Model Higgs boson in $\tau^+\tau^-$ + jets final state, using CDF Run II data with an integrated luminosity of 8.3 fb^{-1}. The Signal considered in this search is four Higgs boson production processes: W/Z Boson associated production ($q\bar{q'} \to WH$, $q\bar{q} \to ZH$) , Vector boson fusion($qq' \to qHq'$) and gluon fusion($gg \to H$), followed by $H \to \tau\tau$.

The τ can decay leptonically – producing an electron or muon – or hadronically. We consider two classes of events as candidates by requiring either one isolated electron/muon and one hadronic τ, or one isolated electron and one isolated muon. We discriminate the Higgs boson signals from backgrounds using a Support Vector Machine which is optimized to maximize search sensitivity. We can not see clear excess of the Higgs boson signal process in kinematic distributions, therefore we extract a 95% confidence level upper limit on the Standard Model Higgs boson cross section($XH \to jj+\tau\tau$). The expected and observed limit are evaluated from $M_H = 100GeV/c^2$ to $M_H = 150GeV/c^2$ in $5GeV/c^2$ step. The observed(expected) limit on assumption of $M_H = 115GeV/c^2$ is $12.2(12.6^{+7.0}_{-3.6})$.

I. INTRODUCTION

One of the most important tasks in particle physics is to observe the standard model(SM) Higgs boson. The existence of the Higgs boson is predicted to explain the mechanism of electroweak symmetry breaking and the origin of mass for the W/Z boson and fermions. Direct searches from the LEP experiment indicate that the SM Higgs boson has to be greater than $114.4GeV/c^2$ at 95% confidence level [1], and indirect constraints (top quark and W boson mass within standard model framework) suggests that it might exist in $M_H = 95GeV^{+34}_{-26}$ at 68% confidence level [2].

At the Tevatron colliding $p\bar{p}$ at $\sqrt{s}$ = 1.96TeV, CDF/D0 experiments perform the search with various SM Higgs boson production and decay modes depending on its mass. In the low mass region ($100GeV < M_H < 135GeV$), the main search mode is W/Z boson associated production with the decay of the SM Higgs boson into a b-quark pair and the W/Z boson into electron or muon. This is because the SM Higgs boson dominantly decays into b-quark pair in this region, and electron or muon from the W/Z boson suppresses QCD background. But these searches are difficult for background modeling and estimation of $b\bar{b}$ process due to poor jet energy resolution compared with leptons(electron/muon). Therefore in low mass region, several subchannels are analyzed to obtain further sensitivity.

Considering this situation, the $H \to \tau\tau$ process is a powerful channel to improve search sensitivity in low mass region. This is because the branching ratio of the SM Higgs boson to a τ pair is the second highest(6.9% at $M_H = 120GeV/c^2$) and four Higgs boson production processes are possible to consider:

- **WH** ($\to \tau\tau$)
- **ZH** ($\to \tau\tau$)
- $\mathbf{qHq}^{(')} \to \mathbf{q}\tau\tau\mathbf{q}^{(')}$
- $\mathbf{gg} \to \mathbf{H} \to \tau\tau$

Also, the main background of this process is an electroweak process, $Z \to \tau\tau$, which has a well estimate production rate and model kinematic shape compared with QCD process. Furthermore, at the LHC, the vector boson fusion production with $H \to \tau\tau$ is one of main channels in low mass range. There is importance to understand the $H \to \tau\tau$ event signature for LHC and Tevatron experiments.

In this analysis, we focus on that final state is τ pair and jets. The additional requirement of one or more jets increases signal acceptance and improve search sensitivity, because we can observe jets with high probability in all signal processes; in the WH/ZH processes, W/Z boson have large branching ratio of decaying into 2jets (67%/70%) and probability of initial state radiation, the vector boson fusion process naturally has jets and gluon fusion process has initial state radiation with high probability. The search of τ pair and leptons(electron or muon) has recently been performed as a new search channel at CDF[3].

We can detect τ only after its decay with weak interaction. One is the leptonic decay ($\tau^{\pm} \to e^{\pm} + \nu_e + \nu_\tau$ / $\tau^{\pm} \to \mu^{\pm} + \nu_\mu + \nu_\tau$) of which B.R is $\sim$ 17% for each lepton type. The other is the hadronic decay ($\tau^{\pm} \to \nu_\tau + h^{\pm}$ / $\tau^{\pm} \to \nu_\tau + h^{\pm}h^{\mp}h^{\pm}$) of which B.R is $\sim$ 65%, where h means charged hadron ($\pi^{\pm}$ or $K^{\pm}$ and so on) and also this decay associate with $\geq$0 neutral hadron (mostly π^0). When 1(3) charged hadron is detected, it called 1(3) prong τ_{had}.

Therefore, in ditau case, there are four mode which could possibly be used: $\tau_{e/\mu}\tau_{e/\mu}$(B.R:6%), $\tau_e\tau_\mu$(B.R:6%), $\tau_{had}\tau_{had}$(B.R:43%) and $\tau_{e/\mu}\tau_{had}$(B.R:46%). Previous analyses at CDF[4][5] considered only e/μ + τ_{had} decay mode which is considered the golden mode because the QCD background is reduced by having one electron or muon, and this mode has the highest branching ratio. We add electron + muon decay mode as new channel to increase signal acceptance and improve sensitivity. On benefit of this channel is that the QCD and Drell-Yan backgrounds are small.

II. THE CDF RUNII DETECTOR

The CDF II Detector is cylindrical and symmetrically surrounds beam pipe by many different layers. Using the azimuthal angle ϕ and the pseudorapidity η, we describe the detector geometry. A complete description of CDF II detector can be found in [6]. Each of these layers work simultaneously with the other components of the detector: Silicon detector, Central Outer Tracker, Solenoid Magnet, Electromagnetic calorimeter, Hadronic calorimeter and Muon detector. The luminosity of $p\bar{p}$ collisions at CDF is measured by Cherenkov luminosity counters.

III. DATA SAMPLE & EVENT SELECTION

This analysis is based on an integrated luminosity of 8.3fb^{-1} collected with CDF II detector from March 2002 thorough March 2011. We applied several cuts to clean up events and these cuts are different between e/μ + τ_{had} channel and e + μ channel. First selections are physical object requirements: in e/μ + τ_{had} channel, we require one electron or muon, one hadronic τ and jets. in e + μ channel, we require one electron, one muon and jets. We further require that electron/muon and hadronic τ or electron and muon have to have opposite charge. This is based on the fact that τ pair should be the opposite sign when those come from the Higgs boson which is neutral particle. In e/μ + τ_{had} channel, we apply additional cut to suppress $Z \rightarrow ee/\mu\mu$ event (Z boson veto). Since the Drell-Yan cross section is relatively large and the final state is actually opposite sign, we try to further reject the $Z \rightarrow ee/\mu\mu$ background. This veto is applied to following events: hadronic τ identified as 1 prong and $E_{had}/P \leq 0.4$, invariant mass of lepton and hadronic τ is between $80GeV/c^2$ and $110GeV/c^2$ (consistent with invariant mass of Z boson). The 0-jet channel is handled as control region and the 1 and $\geq$2 jet channels are used as the signal region for the Higgs boson search.

IV. SIGNAL EXPECTATION

All Signal (WH,ZH,VBF,ggH) Monte Carlo samples are generated by PYTHIA. For each of four Higgs boson production process, there are samples for Higgs boson masses from $100GeV/c^2$ to $150GeV/c^2$ in $5GeV/c^2$ step. The number of expected signal events for each Higgs boson mass point are summarized in Table I. These numbers are after applying detector acceptances and identification efficiencies.

Number of expected signal events						
$\int Ldt = 8.3fb^{-1}$	H $\rightarrow$ $\tau\tau$ + jets					
	e/μ + τ_{had}			e + μ		
$M_H(GeV/c^2)$	= 1jet	$\geq$ 2jet	Total	= 1jet	$\geq$ 2jet	Total
100	6.45	5.15	11.60	0.89	0.67	1.56
105	5.94	4.59	10.52	0.84	0.59	1.43
110	5.30	4.15	9.44	0.73	0.54	1.27
115	4.55	3.54	8.09	0.63	0.46	1.08
120	3.97	3.10	7.07	0.54	0.38	0.92
125	3.33	2.54	5.87	0.44	0.35	0.79
130	2.62	2.06	4.68	0.36	0.27	0.63
135	2.04	1.55	3.59	0.28	0.21	0.45
140	1.47	1.14	2.61	0.21	0.16	0.36
145	1.01	0.77	1.79	0.15	0.10	0.25
150	0.64	0.50	1.13	0.09	0.07	0.16

TABLE I: Number of expected event for each signal channels at each Higgs boson mass point with the number of jet in final state.

V. BACKGROUNDS

Each background process in this analysis is estimated and modeled by Monte Carlo simulations and data-driven techniques. We rely on Monte Carlo (MC) simulations for some background sources: Z/γ^*+jes, $t\bar{t}$, diboson(WW,WZ and ZZ) and a part of W+jets(for kinematic distribution modeling). We use ALPGEN MC samples for Z/γ^*+jes and W+jets. The ALPGEN generator uses PYTHIA for hadronizations and is matched by matrix element and parton shower to avoid double counting[7]. $t\bar{t}$ and diboson(WW/WZ/ZZ) are generated by PYTHIA[8].

The expected number of events is obtained by:

$$N^i = \sigma^i \times Br^i \times A^i \times \epsilon_{trig} \times \epsilon_{ID} \times \epsilon_{vtx} \times \int Ldt, \tag{1}$$

where i denotes each background process. σ^i represents the production cross section for i-th process, and Br^i is the branching ratio for a process i if needed. A^i is the acceptance of i-th process. ϵ_{trig} is the trigger efficiency of *lepton plus track* trigger, ϵ_{ID} is the lepton(including τ_{had}) identification and reconstruction scale factor, and ϵ_{vtx} is the run dependent efficiency of z vertex position requirement. $\int Ldt$ is the integrated luminosity used this analysis.

As for data driven techniques, the same sign (SS) data is directly used for the estimation and modeling of QCD, γ+jet and W+jets, these can contribute into the signal region when jets fake $e/\mu/\tau_{had}$. In order to use the same sign data, there is one assumption for QCD and γ+jets processes: in final state of these processes, there is no charge correlation which is expected between the charge of reconstructed e/μ and τ_{had} or e and μ. Stated another way, the number of opposite sign (OS) events should be equal to the number of same sign(SS) events. In W+jets process, we can also use same sign data for estimating and modeling W+jets events, but this process has a charge correlation in e/μ from W boson and the outgoing quark. There are more opposite sign events than same sign events($N_{OS} \geq N_{SS}$).

In e + μ channel, this additional contribution is very small which can be ignored because $jet \rightarrow e/\mu$ fake rate is not very high. On the other hand in e/μ + τ_{had} channel, $jet \rightarrow \tau_{had}$ fake rate($\sim$10%) is higher than $jet \rightarrow e/\mu$. Therefore, we should not directly use same sign data for W+jets estimation without estimating additional contributions to this background, which means $N_{add-on} = N_{OS} - N_{SS}$.

In order to perform this estimation, we study the W+jets control region. We require one loose τ_{had} which is defined by lowering thresholds of some ID cuts (seed track P_T, visible P_T and visible Mass) and exclusive τ_{had} ID used in the Higgs boson signal selection. Also we require one electron or muon, and apply a missing transverse energy($\not\!\!E_T$) cut ($\not\!\!E_T \leq 25/30/35 GeV$ for 0/1/$\geq$2 jet), and the transverse mass of e/μ and $\not\!\!E_T$ cut ($M_T(e/\mu, \not\!\!E_T) > 40 GeV/c^2$). Using this region, we evaluate a scale factor between the number of opposite sign events and the number of same sign events. Then the number of additional W+jets events are estimated using this scale factor. Kinematical distributions for additional W+jets events are represented by ALPGEN MC samples.

After these estimations, the expected number of background and observed data are summarized in Table II and Table III. The prediction and observed data are consistent within uncertainties.

H $\rightarrow$ $\tau\tau$ + jets (e/μ + τ_{had} channel)			
Source	0 jet	1 jet	$\geq$2 jet
$Z \rightarrow \tau\tau$	12390.6 ± 1078.0	2471.9 ± 311.5	609.1 ± 100.5
$Z \rightarrow ee/\mu\mu$	763.2 ± 77.1	199.6 ± 25.5	51.9 ± 7.8
$t\bar{t}$	0.5 ± 0.2	12.3 ± 2.4	138.3 ± 18.0
WW/WZ/ZZ	109.9 ± 12.1	96.2 ± 9.7	45.7 ± 6.7
jet$\rightarrow \tau_{had}$	19331.0 ± 139.0	5843.0 ± 128.0	1551 ± 39.4
Add-on. W+jets	2393.3 ± 363.8	785.0 ± 76.4	169.9 ± 41.5
Total Background	34988.5 ± 1148.8	9408.0 ± 346.4	2566.9 ± 115.8
$H \rightarrow \tau\tau(M_H = 120GeV/c^2)$	–	3.97 ± 0.75	3.10 ± 1.12
Observed $(8.3fb^{-1})$	34983	9278	2633

TABLE II: Background prediction and observed data for e/μ + τ_{had} channel with 8.3fb^{-1}.

H $\rightarrow$ $\tau\tau$ + jets (e + μ channel)			
Source	0 jet	1 jet	$\geq$2 jet
$Z \rightarrow \tau\tau$	1384.9 ± 113.6	283.6 ± 32.6	74.5 ± 14.0
$Z \rightarrow ee/\mu\mu$	83.4 ± 8.2	29.6 ± 3.5	11.4 ± 1.8
$t\bar{t}$	0.6 ± 0.2	17.0 ± 3.0	126.5 ± 16.1
WW/WZ/ZZ	160.3 ± 16.2	43.5 ± 5.1	13.7 ± 2.3
jet$\rightarrow e/\mu$	187.0 ± 13.7	54.0 ± 7.3	20.0 ± 4.5
Total Background	1816.2 ± 115.8	427.7 ± 34.1	246.1 ± 22.0
$H \rightarrow \tau\tau(M_H = 120GeV/c^2)$	–	0.53 ± 0.09	0.38 ± 0.16
Observed $(8.3fb^{-1})$	1811	408	260

TABLE III: Background prediction and observed data for e + μ channel with 8.3fb^{-1}.

VI. DISCRIMINATION

To improve sensitivity, a Support Vector Machine (SVM) [10] is trained using several kinematical variables to discriminate $H \rightarrow \tau\tau$ from backgrounds. Support Vector Machines are a multivariate analysis method. The concept of SVM is very simple: use a decision plane to separate signal and background in hyper space(kinematical variables space). We use SVM in TMVA tool kit (TMVA v4.1.0 and ROOT v5.28).

Four signal channels are defined in this analysis because these background components are respectively different.

- **e/μ + τ_{had} + 1jet channel**
- **e/μ + τ_{had} + $\geq$ 2jet channel**
- **e + μ + 1jet channel**
- **e + μ + $\geq$ 2jet channel**

We prepare four classifiers for each 1jet and $\geq$2jet in e/μ + τ_{had} channel and two classifiers for 1jet and $\geq$ 2jet in e + μ channel. Each classifier is trained using different input variables. The H$\rightarrow$ $\tau\tau$ process is defined as signal and one of the background processes is defined as background, which is dominant in background source or has different kinematical distributions from signal. The definition of each classifier is summarized in Table IV.

H $\rightarrow$ $\tau\tau$ + jets			
e/μ + τ_{had} channel		e + μ + channel	
1 jet	$\geq$ 2jet	1 jet	$\geq$ 2jet
$H \rightarrow \tau\tau$ vs. Z$\rightarrow \tau\tau$	$H \rightarrow \tau\tau$ vs. Z$\rightarrow \tau\tau$	$H \rightarrow \tau\tau$ vs. Z$\rightarrow \tau\tau$	$H \rightarrow \tau\tau$ vs. Z$\rightarrow \tau\tau$
$H \rightarrow \tau\tau$ vs. QCD	$H \rightarrow \tau\tau$ vs. QCD	$H \rightarrow \tau\tau$ vs. QCD	$H \rightarrow \tau\tau$ vs. QCD
$H \rightarrow \tau\tau$ vs. W+jets	$H \rightarrow \tau\tau$ vs. W+jets		
$H \rightarrow \tau\tau$ vs. diboson	$H \rightarrow \tau\tau$ vs. $t\bar{t}$		

TABLE IV: The training category for each signal channel.

For the signal training sample, we use PYTHIA MC sample for each process: WH$\rightarrow 2\tau + 2jets$, $ZH \rightarrow 2\tau + 2jets$, $VBF \rightarrow 2\tau + 2jets$, $ggH \rightarrow 2\tau(+jets)$. Each classifier is trained separately by each Higgs boson mass point: $110GeV/c^2$, $115GeV/c^2$, $120GeV/c^2$, $130GeV/c^2$, $140GeV/c^2$, $150GeV/c^2$. Each signal sample is mixed by weighting the number of expected events at each mass point.

For the background training samples, ALPGEN MC samples are used for $Z/\gamma* \rightarrow \tau\tau$, W+jets and PYTHIA MC samples are used for $t\bar{t}$ and diboson. There are three independent processes in diboson process: WW, WZ, ZZ. These processes are mixed by weighting the expected number of events for the diboson training sample. For the QCD training sample, we use data which are in isolation-side-band region. This means that e/μ isolation is more than 0.4, that these events are not in our candidate events.

Finally, we obtain four classifiers for each e/μ + τ_{had} channel and two classifiers for e + μ channel. Final discriminant distribution is obtained by taking MINIMUM response among distributions from classifiers. Figure 1 shows final discriminant distributions for e/τ_{had} + 1/$\geq$2jet channels and e + μ + 1/$\geq$ 2 jet channels. In these figures, signal distribution is scaled by one hundred, and assumed that $M_H = 115GeV/c^2$.

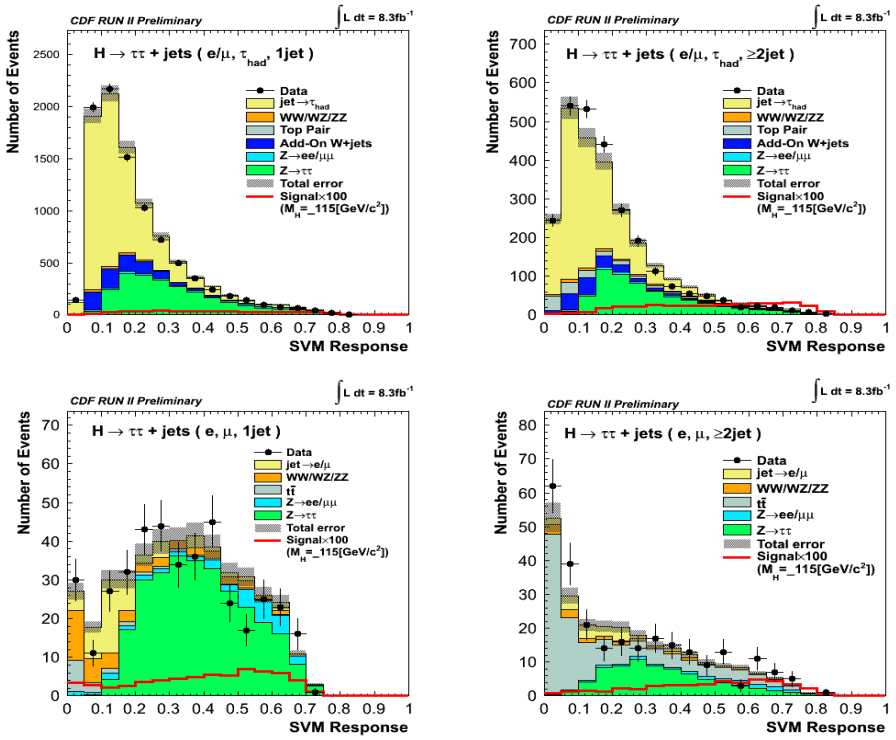

FIG. 1: Final discriminant distributions for each channel at $M_H = 115GeV/c^2$

VII. SYSTEMATIC UNCERTAINTIES

Several sources of systematic uncertainties are taken into account in this analysis. We assign 5.9% uncertainty for the integrated luminosity and 0.5% uncertainty for the efficiency of z vertex position cut. The trigger efficiency and lepton identification and reconstruction scale factor uncertainties depend on lepton type.

We assign 0.3% , 1.0% and 3.0% as trigger efficiency uncertainties for electron, muon and τ_{had}, and also assign 2.4%, 2.6% and 3.0% as identification and reconstruction scale factor uncertainties. Uncertainties for theoretical cross section depend on the process: For Drell-Yan processes, 2.2% uncertainty is assigned for all lepton types based on previous CDF measurement uncertainty[9]. For $t\bar{t}$ and diboson processes, 6% and 10% uncertainties are assigned respectively. For four Higgs boson production processes these uncertainties are considered: 5% for W/Z boson associated production, 10% for vector boson fusion and 23.5%(33.0%) for gluon fusion in the 1jet($\geq$2jet) channel.

For the normalization factor of additional W+jets contribution, since these values are obtained completely from data, we do not assign any uncertainty related to Monte Carlo simulation, but only assign related to data/MC scale factor uncertainty of 14.8% (1.4 ± 0.2). The uncertainty of same sign data method are assigned statistic error. Uncertainties related to parton distribution function are taken into account for all Higgs boson signal processes. We produced weight functions in CTEQ6M and also for MRST72 and for MRST75. Then these two are added in quadrature. Uncertainties related to initial state radiation (ISR) and final state radiation (FSR) are take into account for all Higgs boson signal processes. These values are evaluated for each signal channel, by calculating acceptance difference in MC samples, which is increased or reduced the strength of ISR and FSR in $\pm 1\sigma$ at the parton showering process.

The main source of systematic uncertainty is related to jet energy scale. The effect of jet energy scale is evaluated by applying jet energy correction shifted $\pm 1\sigma$ from default correction factor. The uncertainty in the shape of final discriminant due to jet energy scale is also considered. These systematic uncertainties are summarized in Table V and Table VI

Systematic Uncertainties (%)											
$H \to \tau\tau$ + jets (e/μ + τ_{had} channel)											
Source		$Z \to \tau\tau$	$Z \to ll$	$t\bar{t}$	diboson	$jet \to \tau_{had}$	Add.Wjet	WH	ZH	VBF	ggH
Luminosity		5.9	5.9	5.9	5.9	-	-	5.9	5.9	5.9	5.9
$\lvert Z_{vertex}\rvert$		0.5	0.5	0.5	0.5	-	-	0.5	0.5	0.5	0.5
σ/Norm.	$= 1jet$	2.0	5.0	10.0	6.0	1.3	14.8	5.0	5.0	10.0	23.5
	$\geq 2jet$	2.0	5.0	10.0	6.0	2.5	14.8	5.0	5.0	10.0	33.0
PDF		-	-	-	-	-	-	1.2	0.9	2.2	4.9
ele Trigger		0.3	0.3	0.3	0.3	-	-	0.3	0.3	0.3	0.3
μ Trigger		1.0	1.0	1.0	1.0	-	-	1.0	1.0	1.0	1.0
track Trigger		3.0	3.0	3.0	3.0	-	-	3.0	3.0	3.0	3.0
ele ID		2.4	2.4	2.4	2.4	-	-	2.4	2.4	2.4	2.4
μ ID		2.6	2.6	2.6	2.6	-	-	2.6	2.6	2.6	2.6
τ_{had} ID		3.0	3.0	3.0	3.0	-	-	3.0	3.0	3.0	3.0
JES	$= 1jet$	9.5	8.5	14.5	0.5	-	4.2	2.8	6.4	6.5	4.3
	$\geq 2jet$	14.2	12.1	1.3	10.7	-	15.4	5.1	3.9	3.7	14.5
IF-SR	$= 1jet$	-	-	-	-	-	-	6.7	8.7	8.8	3.6
	$\geq 2jet$	-	-	-	-	-	-	4.8	3.8	3.9	19.1
Total	$= 1jet$	12.6	12.8	19.4	10.1	1.3	16.3	12.1	14.5	17.1	26.0
	$\geq 2jet$	16.5	14.7	12.9	14.7	2.5	21.4	11.9	11.1	14.2	41.9

TABLE V: Summary of systematic uncertainties in e/μ + τ_{had} channel.

Systematic Uncertainties (%)										
H $\to \tau\tau$ + jets (e + μ channel)										
Source		$Z \to \tau\tau$	$Z \to ll$	$t\bar{t}$	diboson	$jet \to e/\mu$	WH	ZH	VBF	ggH
Luminosity		5.9	5.9	5.9	5.9	-	5.9	5.9	5.9	5.9
$\lvert Z_{vertex}\rvert$		0.5	0.5	0.5	0.5	-	0.5	0.5	0.5	0.5
σ/Norm	$= 1jet$	2.0	5.0	10.0	6.0	13.7	5.0	5.0	10.0	23.5
	$\geq 2jet$	2.0	5.0	10.0	6.0	22.3	5.0	5.0	10.0	33.0
PDF		-	-	-	-	-	1.2	0.9	2.2	4.9
ele Trigger		0.3	0.3	0.3	0.3	-	0.3	0.3	0.3	0.3
μ Trigger		1.0	1.0	1.0	1.0	-	1.0	1.0	1.0	1.0
track Trigger		3.0	3.0	3.0	3.0	-	3.0	3.0	3.0	3.0
ele ID		2.4	2.4	2.4	2.4	-	2.4	2.4	2.4	2.4
μ ID		2.6	2.6	2.6	2.6	-	2.6	2.6	2.6	2.6
JES	$= 1jet$	8.3	7.7	12.3	6.7	-	6.7	8.7	8.8	3.6
	$\geq 2jet$	16.8	11.0	1.8	13.6	-	4.8	3.8	3.9	19.1
IF-SR	$= 1jet$	-	-	-	-	-	6.4	9.9	4.0	7.6
	$\geq 2jet$	-	-	-	-	-	3.4	1.1	0.8	31.1
Total	$= 1jet$	11.5	11.9	17.6	11.8	13.7	11.3	14.0	15.5	26.5
	$\geq 2jet$	18.6	14.3	12.7	16.7	22.3	11.7	10.2	13.4	49.4

TABLE VI: Summary of systematic uncertainties in e + μ channel.

VIII. RESULTS

From the expected sensitivity, it is already known to be quite hard to claim an evidence of the SM Higgs boson and in fact there is no clear excess in the signal region. Given the fact that our background models nicely agree with data in background rich region, we extract 95% C.L limit for the SM Higgs boson cross section. Using final discriminant distributions, we perform binned maximum likelihood method to obtain 95% C.L limit. For each bin, the expected number of event(μ_i) is evaluated as below:

$$\mu_i = \sum_{k=1}^{N_{bkg}} f_i^k \cdot N^k + \sum_{k=1}^{N_{sig}} f_i^l \cdot \left(\epsilon^l \cdot \sigma^l \cdot \int Ldt \right), \tag{2}$$

where k represents the kind of background($Z \to \tau\tau$, $Z \to ll$, $t\bar{t}$, diboson, add-on W+jets) and l represents the kind of signal(WH, ZH, VBF, ggH), and N_{bng} and N_{sig} are the number of kind of background process($N_{bkg} = 5$) and signal process($N_{sig} = 4$). For each k and l, f_i^k and f_i^l represent the expected fraction in i-th bin. In the second term, ϵ^l is the detection efficiency (Acceptance, trigger efficiency, lepton identification efficiency, z vertex position cut efficiency) for signal l and $\int Ldt$ is the integrated luminosity. And finally, σ^l is production cross section of signal l, which is unknown parameter for likelihood fit.

Then, we define likelihood function as below:

$$L\left(\frac{\sigma}{\sigma_{SM}}\right) = \int \cdots \int \prod_{i=1}^{N_{bin}} \frac{\mu_i^{N_i}}{N_i!} e^{-\mu_i} \prod_{k=1}^{N_{bkg}} G\left(N^k, \Delta^k dN^k\right) \prod_{l=1}^{N_{sig}} G\left(N^l, \Delta^l dN^l\right), \tag{3}$$

where Δ^k and Δ^l are the systematic uncertainty of each processes, and G represents gaussian fraction, which is fluctuated by the expected systematic uncertainty. N_i is the number of observed event in i-th bin. For each signal production cross section(σ^l), we assume that these are 100% correlated. Therefore ratios between parameters of σ^l and the SM production cross section(σ_{SM}) are same, we consider one parameter ($\frac{\sigma}{\sigma_{SM}}$) in the likelihood.

Here, we define a likelihood for each signal channel. Then, four likelihoods are collected by taking

their product:

$$L = L_1 \times L_2 \times L_3 \times L_4. \tag{4}$$

We evaluate the expected 95% C.L limit from the maximum binned likelihood by pseudo-experiment. To get the number of events of pseudo data, each background source is fluctuated by gaussian fraction with systematic uncertainties. Then, the number of events for each bin in one pseudo-experiment is extracted by Poisson distribution. Finally, we summarize the combined limits from the four channels in Table VII and Figure 2 and 3 as the result of this analysis.

H $\rightarrow$ $\tau\tau$ + jets						
$\int Ldt = 8.3fb^{-1}$	Expected Limit/σ_{SM}					Observed Limit/σ_{SM}
$M_H(GeV/c^2)$	-2σ	-1σ	Median	+1σ	+2σ	
100.0	7.4	11.3	16.8	25.8	39.3	12.5
105.0	6.5	9.6	14.9	23.4	34.9	11.2
110.0	6.2	9.0	13.1	20.2	28.8	10.4
115.0	5.9	8.9	12.6	19.6	29.9	12.2
120.0	6.8	9.4	13.9	20.4	30.7	11.1
125.0	6.9	10.2	14.8	21.8	33.0	11.7
130.0	8.0	10.8	16.1	24.2	34.5	12.4
135.0	10.4	14.6	21.0	30.2	44.1	23.6
140.0	11.9	17.5	25.1	36.6	51.9	27.5
145.0	16.7	22.9	33.0	48.2	68.6	29.7
150.0	23.0	33.5	47.5	69.1	90.2	43.0

TABLE VII: Expected and Observed limit at 95% confidence level

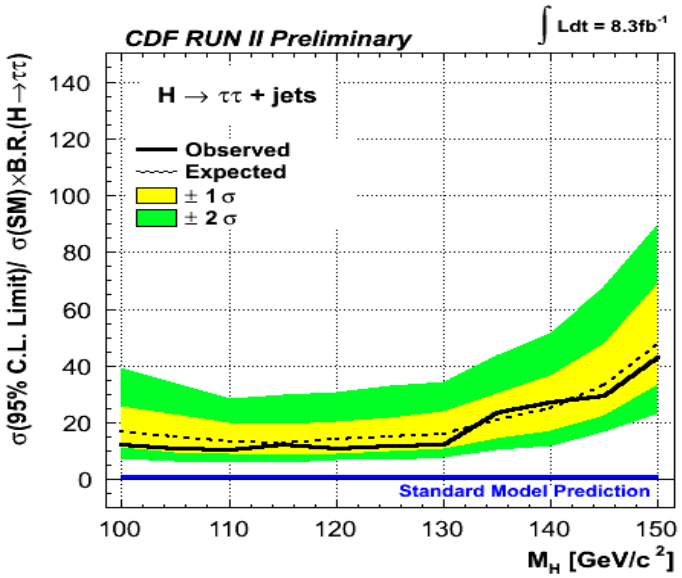

FIG. 2: Expected and Observed limit at 95% confidence level (linear scale)

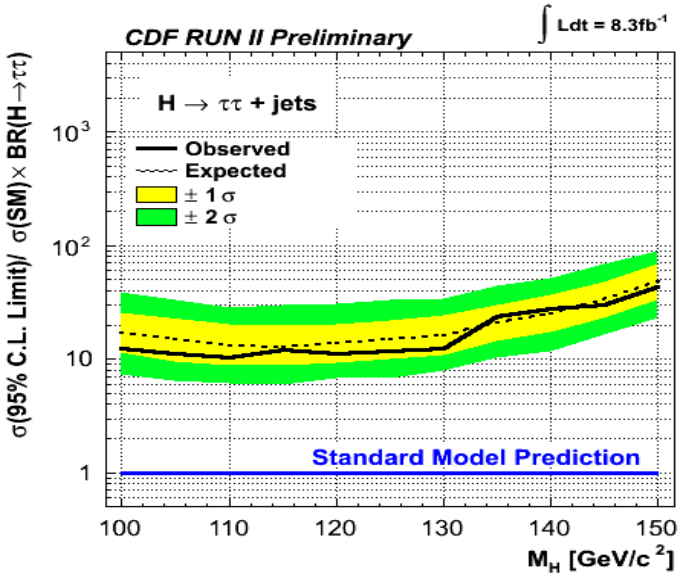

FIG. 3: Expected and Observed limit at 95% confidence level (log scale)

IX. CONCLUSION

We performed a search for the standard model Higgs boson in $\tau\tau$ + jets final state with 8.3fb^{-1} collected by CDF II detector. Since there was no significant excess in kinematical distributions, we extract a cross section upper limit on the standard model Higgs boson($XH \rightarrow \tau\tau + jets$) at 95% confidence level. The observed(expected) cross section limit for $M_H = 115GeV/c^2$ is 12.2(12.6$^{+7.0}_{-3.6}$).

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

[1] G. Abbiendi, *et al.* (the ALEPH Collaboration, the DELPHI Collaboration, the L3 Collaboration and the OPAL Collaboration, The LEP Working Group for Higgs Boson Searches), Search for the Standard Model Higgs Boson at LEP, arXiv:hep-ex/0306033v1 (2003)

[2] The LEP Electroweak Working Group, http://lepewwg.web.cern.ch/LEPEWWG/

[3] Koji Ebina, Kohei Yorita, Search for the Standard Model Higgs in $\ell\nu + \tau\tau$ and $\ell\ell + \tau\tau$ channel, CDF public note 10500
[4] Kohei Yorita, Yound-Kee Kim, Search for Standard Model Higgs Boson in H-¿tautau Channel with 2fb^{-1}, CDF Public note 9248
[5] Pierluigi Totaro, Donatella Lucchesi, Anna Maria Zanetti, Melisa Rossi, Search for a low mass Standard Model Higgs boson in the di-tau decay channel using 6.0 fb^{-1}, CDF public note 10439
[6] R. Blair *et al.*, The CDF Collaboration, FERMILAB-PUB-96/390-E (1996).
[7] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, ALPGEN, a generator for hard multi-parton processes in hadronic collisions, JHEP 0307 (2003) 001
[8] T. Sjostrand *et al.*, High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. **135**, 238 (2001).
[9] D. Acosta *et al.*, Phys. Rev.Lett. 94, 091803 (2005)
[10] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, "TMVA 4 Toolkit for Multivariate Data Analysis with ROOT Users Guide", arXiv:physics/0703039a
C. Cortes and V. Vapnik, "Support vector networks", Machine Learning, 20, 273 (1995).
V. Vapnik, "The Nature of Statistical Learning Theory", Springer Verlag, New York, 1995.
C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2, 1 (1998).